

RESEARCH ARTICLE

10.1002/2015JD024715

Surrogate gas prediction model as a proxy for $\Delta^{14}\text{C}$ -based measurements of fossil fuel CO_2 Kevin J Coakley¹, John B Miller^{2,3}, Stephen A Montzka², Colm Sweeney^{2,3}, and Ben R Miller^{2,3}¹Statistical Engineering Division, National Institute of Standards and Technology, Boulder, Colorado, USA, ²Global Monitoring Division, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA, ³CIRES, University of Colorado Boulder, Boulder, Colorado, USA

Key Points:

- Proxy-gas model for fossil fuel CO_2 measurements presented
- Method may increase number of fossil fuel- CO_2 estimates by up to three times
- Uncertainties of proxy model predictions quantified

Supporting Information:

- Supporting Information S1
- Text S1
- Text S2
- Text S3
- Text S4
- Data Set S1
- Data Set S2
- Data Set S3
- Data Set S4
- Data Set S5
- Data Set S6
- Data Set S7
- Data Set S8

Correspondence to:

K. J. Coakley,
kevin.coakley@nist.gov

Citation:

Coakley, K. J., J. B. Miller, S. A. Montzka, C. Sweeney, and B. R. Miller (2016), Surrogate gas prediction model as a proxy for $\Delta^{14}\text{C}$ -based measurements of fossil fuel CO_2 , *J. Geophys. Res. Atmos.*, 121, 7489–7505, doi:10.1002/2015JD024715.

Received 29 DEC 2015

Accepted 25 MAY 2016

Accepted article online 2 JUN 2016

Published online 27 JUN 2016

Abstract The measured $^{14}\text{C}:^{12}\text{C}$ isotopic ratio of atmospheric CO_2 (and its associated derived $\Delta^{14}\text{C}$ value) is an ideal tracer for determination of the fossil fuel derived CO_2 enhancement contributing to any atmospheric CO_2 measurement (C_{ff}). Given enough such measurements, independent top-down estimation of U.S. fossil fuel CO_2 emissions should be possible. However, the number of $\Delta^{14}\text{C}$ measurements is presently constrained by cost, available sample volume, and availability of mass spectrometer measurement facilities. $\Delta^{14}\text{C}$ is therefore measured in just a small fraction of samples obtained by flask air sampling networks around the world. Here we develop a projection pursuit regression (PPR) model to predict C_{ff} as a function of multiple surrogate gases acquired within the NOAA/Earth System Research Laboratory (ESRL) Global Greenhouse Gas Reference Network (GGGRN). The surrogates consist of measured enhancements of various anthropogenic trace gases, including CO , SF_6 , and halocarbon and hydrocarbon acquired in vertical airborne sampling profiles near Cape May, NJ and Portsmouth, NH from 2005 to 2010. Model performance for these sites is quantified based on predicted values corresponding to test data excluded from the model building process. Chi-square hypothesis test analysis indicates that these predictions and corresponding observations are consistent given our uncertainty budget which accounts for random effects and one particular systematic effect. However, quantification of the combined uncertainty of the prediction due to all relevant systematic effects is difficult because of the limited range of the observations and their relatively high fractional uncertainties at the sampling sites considered here. To account for the possibility of additional systematic effects, we incorporate another component of uncertainty into our budget. Expanding the number of $\Delta^{14}\text{C}$ measurements in the NOAA GGGRN and building new PPR models at additional sites would improve our understanding of uncertainties and potentially increase the number of C_{ff} estimates by approximately a factor of 3. Provided that these estimates are of comparable quality to $\Delta^{14}\text{C}$ -based estimates, we expect an improved determination of fossil fuel CO_2 emissions.

1. Introduction

Over the last 200 years, the amount fraction of CO_2 in the atmosphere has increased from approximately 280 μmol of CO_2 per mole of dry air to nearly 400 μmol of CO_2 per mole of dry air [Ballantyne *et al.*, 2012; Etheridge *et al.*, 1996]. (Henceforth, for the sake of brevity, we denote micromole per mole as ppm.) There is an overwhelming scientific consensus that this increase is due primarily to fossil fuel emissions despite the fact that about half of the CO_2 produced by burning fossil fuels is absorbed by oceans and the terrestrial biosphere [Canadell *et al.*, 2007; Knorr, 2009; Ballantyne *et al.*, 2012]. Although year-to-year increases in CO_2 are well explained by the global average of remote atmosphere measurements, variability at shorter time scales over large land areas is often dominated by terrestrial biosphere exchange. To fully understand the existing array of atmospheric CO_2 measurements in terms of potential sources and sinks, determinations of CO_2 enhancements due solely to combustion of fossil fuels, C_{ff} , are essential. Currently, C_{ff} is estimated from inventories of fossil fuel CO_2 emissions that are based on economic statistics [Gurney *et al.*, 2009; Boden *et al.*, 2014]. Atmosphere-based determinations of a sufficiently large number of C_{ff} values should enable both near real-time monitoring of fossil fuel emissions and an independent assessment of bottom-up inventories.

Due to nuclear reactions associated with cosmic rays, ^{14}C is continually produced in the atmosphere and then rapidly oxidized to produce $^{14}\text{CO}_2$ [Anderson *et al.*, 1947; Lingenfelter, 1963; Suess, 1965]. Atmospheric testing

of nuclear weapons during the 1960s nearly doubled the amount of $^{14}\text{CO}_2$ in the Northern Hemisphere atmosphere at one point in time [Lal and Suess, 1968]. Since then, much of this excess has been absorbed by the oceans and biosphere. Nuclear reactors also produce $^{14}\text{CO}_2$ [Yim and Caron, 2006] at spatially varying rates. Notably, these rates are generally higher in Europe than in the U.S. [Graven and Gruber, 2011; Vogel et al., 2013]. Although the effect of nuclear reactor emissions on C_{ff} has been assessed at CMA (Cape May, NJ) and NHA (Portsmouth, NH) [Lehman et al., 2013], in order to develop proxy prediction models elsewhere, one must assess nuclear reactor effects on a site-to-site basis.

Since underground fossil fuels stored in the rock reservoir were last in contact with the atmosphere hundreds of millions years ago, and $^{14}\text{CO}_2$ decays with a half-life of (5700 ± 30) years (National Nuclear Data Center, Brookhaven National Laboratory, www.nndc.bnl.gov), the isotopic ratio $^{14}\text{C}:^{12}\text{C}$ of CO_2 produced by burning fossil fuels is negligible and equated to zero in any analysis. To understand this point more clearly, note that the probability that a ^{14}C nucleus does not decay until after 100×10^{-6} years is 10^{-5281} given that its half-life is 5700 years. Since CO_2 contributed by nearly all other sources is in near equilibrium with the atmosphere, a measurement of $^{14}\text{C}:^{12}\text{C}$ and its associated derived $\Delta^{14}\text{C}$ value [Stuiver and Pollack, 1977] (note that $\Delta^{14}\text{C}$ corresponds to Δ in Stuiver and Pollack [1977]) enables one to determine C_{ff} according to relatively simple mass balance considerations [Levin et al., 2003; Levin and Karstens, 2007; Turnbull et al., 2007, 2011; Vogel et al., 2010; Miller et al., 2012]. Approximately 5000 high precision measurements of $\Delta^{14}\text{C}$ (with random uncertainty approximately 0.2%) and associated derived C_{ff} estimates (with random uncertainty approximately 1 ppm) would enable an independent “top-down” (i.e., atmosphere-based) estimation of monthly averages of U.S. national and regional fossil fuel CO_2 emissions with random uncertainties in the 5%–10% range [Committee on Methods for Estimating Greenhouse Gas Emissions; National Research Council, 2010; Basu et al., 2016]. However, our ability to make the necessary $\Delta^{14}\text{C}$ measurements (e.g., several thousand per year just for the United States) is currently constrained by cost, accessibility to accelerator mass spectrometers (AMS), and the volume of air (approximately 2 L at standard temperature and pressure (273.15 K and 101.325 kPa) [Lehman et al., 2013]) required to make high precision measurements of $\Delta^{14}\text{C}$. Thus, $\Delta^{14}\text{C}$ is currently measured in just a small subset of air samples obtained from sampling networks (including the U.S. portion of the NOAA/ESRL Global Greenhouse Gas Reference Network (GGGRN)). A substantial number of new C_{ff} measurements determined from either $\Delta^{14}\text{C}$ or surrogate gases measurements would greatly facilitate efforts to verify “bottom-up” inventory approaches for U.S. fossil fuel CO_2 emissions.

Various studies have presented univariate proxy models for C_{ff} based on CO [Turnbull et al., 2011; Vogel et al., 2010]. In these studies, C_{ff} is predicted based on interpolated measured ratios of $\text{CO}:C_{\text{ff}}$. However, spatiotemporal extrapolation of measured ratios of $\text{CO}:C_{\text{ff}}$ is problematic because the relative intensity of nonfossil fuel CO sources can vary as a complicated function of time and space. Further, CO-based proxies can be problematic because many sources of fossil fuel emissions such as vehicles with diesel engines and power plants have very low CO emission rates. Miller et al. [2012] demonstrated that a wide range of other anthropogenic trace gases such as SF_6 , and a number of halocarbon and hydrocarbon that are frequently measured in NOAA network samples are positively correlated with observed C_{ff} . Because of these correlations, a multivariate proxy prediction model for C_{ff} based on measured surrogate gases is plausible. If one could accurately predict C_{ff} as a function of multiple anthropogenic trace gases over broad spatiotemporal regions of the U.S., such a prediction model could serve as proxy for a $\Delta^{14}\text{C}$ -based measurement of C_{ff} and expand the temporal density of reliable C_{ff} estimates ($\Delta^{14}\text{C}$ -based measurements and proxy model predictions) by a factor of approximately 3 at towers and aircraft vertical profiling sites throughout North America [Sweeney et al., 2015; Andrews et al., 2014]. In this work, we construct such a prediction model based on multiple surrogate gas measurements and quantify its performance. To the best of our knowledge, our model is the first multivariate proxy prediction model for C_{ff} .

2. Methods

We develop projection pursuit regression (PPR) [Friedman and Stuetzle, 1981] prediction models based on $\Delta^{14}\text{C}$ -derived measurements of C_{ff} and measured enhancements of twelve anthropogenic surrogate gases acquired by NOAA from vertical aircraft profiles at two sites, CMA and NHA (see Figure 1 and Table 1), between 2005 and 2010. A large fraction of these measurements (between 2005 and 2009) were first presented and analyzed in Miller et al. [2012]. The rest of the measurements were acquired between 2009 and 2010.

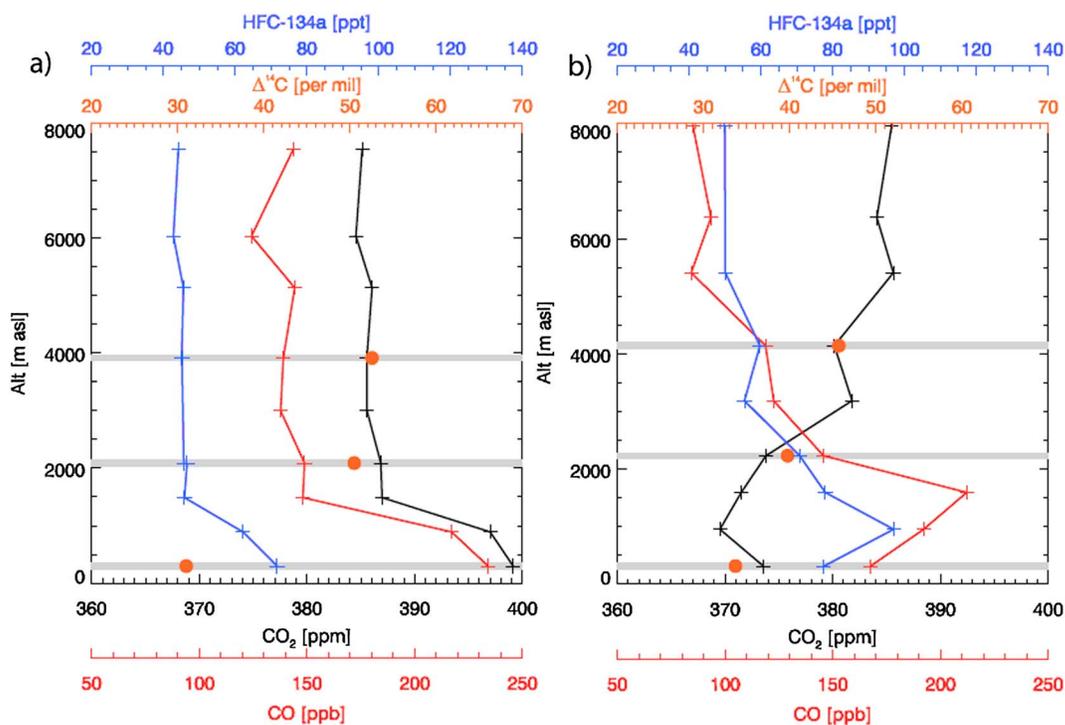


Figure 1. Example vertical profiles (a) above site CMA (offshore from Cape May, NJ) from 21 February 2007 and (b) above site NHA (offshore from Portsmouth, NH) from 10 July 2008. Black, red, and blue pluses connected with a line represent nine air samples collected between the surface and 8 km asl, with measured mole fractions shown for CO₂, CO, and HFC-134a, respectively. Red circles represent values from the three typical altitudes where Δ¹⁴C is analyzed. Grey bars highlight the nominal altitudes at which all gases were measured (reprinted figure from Miller et al. [2012]).

Given these C_{ff} and surrogate gas data, we determine which trace gases to include in PPR prediction models, as well as their complexity and form, with a statistical learning method [Hastie et al., 2008] called cross validation [Arlot and Celisse, 2010; Stone, 1974, 1977]. Following Hastie et al. [2008], we split the observed data into model building data and test data. The model building data subset is then split into training data and validation data subsets. After fitting each candidate model to the training data, we predict C_{ff} values corresponding to observed values in the validation data. We select the model that minimizes the discrepancy between the observed and predicted C_{ff} values in the validation data. Since both the training and validation data are involved in the model selection process, quantification of prediction model performance based on how well

the model fits the validation data is an imperfect and usually overly optimistic measure of how well the model performs for new data excluded from the model building process. To better quantify model performance for new data, we predict C_{ff} corresponding to test data excluded from the model building process.

We stress that the only inputs to the PPR prediction model are measured surrogate gas enhancements. That is, the time and location of any surrogate gas enhancement measurement are not predictors in the PPR model. Later in this work, we discuss how possible spatial and temporal effects such as systematic temporal variations of surrogate gas emission rates could affect the performance of our prediction model. Next, we present measurement models for C_{ff} and surrogate gas enhancements following Miller et al. [2012].

Table 1. Surrogate Gases

Trace Gas index	name
1	CO
2	SF ₆
3	HFC-134a
4	HCFC-22
5	HCFC-142b
6	C ₂ Cl ₄
7	CH ₂ Cl ₂
8	HFC-152a
9	C ₆ H ₆
10	CH ₄
11	N ₂ O
12	CFC-11

2.1. Measurement Models and Data

The estimated value $C_{ff,obs}$ of the theoretical (true) value C_{ff} for any sample is based, in part, on the measured isotopic ratio of ^{14}C and ^{12}C , $^{14}\text{R}_{obs} = ^{14}\text{C}/^{12}\text{C}$, for that sample. Following *Stuiver and Pollack* [1977], we estimate the theoretical value of $\Delta^{14}\text{C}$ as Δ_{obs} , where

$$\Delta_{obs} = ^{14}\text{R}_{obs}/^{14}\text{R}_{standard} \times \Theta_{obs} - 1, \quad (1)$$

and Θ_{obs} is a correction term (which depends on the measured isotopic ratio of ^{13}C and ^{12}C for the sample) that accounts for systematic effects due to mass-dependent fractionation. Above, isotopic ratio estimates are also corrected for small radioactive decay losses during the interval between acquisition and measurement of the sample.

We model the theoretical values of C_{ff} and surrogate gas enhancements with the simple one-dimensional analytical framework employed by *Miller et al.* [2012]. In this approach, the theoretical mole fraction of a gas at lower levels (typically at approximate altitudes of 300 m and 2200 m above sea level (asl)) is the sum of a theoretical background plus a theoretical enhancement due to recent anthropogenic emissions of that gas at the site of interest. Further, we assume that the theoretical background at lower levels is the same as the theoretical background at upper levels in the relatively well-mixed free troposphere (typically 4000 m asl) and that theoretical enhancements are 0 at upper levels. Based on these assumptions, we estimate any particular surrogate gas enhancement (i.e., an additive signal relative to a background) as the difference between measured mole fractions at lower and upper levels of a particular sampling profile.

Following earlier work [*Turnbull et al.*, 2007; *Miller et al.*, 2012], we estimate C_{ff} as $C_{ff,obs}$, where

$$C_{ff,obs} = \frac{C_{obs}(\Delta_{obs} - \Delta_{bg,obs})}{\Delta_{ff} - \Delta_{bg,obs}} - \hat{C}_{cor}, \quad (2)$$

where $\Delta_{ff} = -1$, and the measured values C_{obs} , Δ_{obs} , $\Delta_{bg,obs}$ are estimates of the unknown theoretical values C , Δ and Δ_{bg} (see Appendix A for more details). Above, Δ is shorthand for measured $\Delta^{14}\text{C}$, C is shorthand for CO_2 , and \hat{C}_{cor} is an estimate of small contributions to the tropospheric ^{14}C budget that may influence the measured enrichment or depletion relative to background. These contributions include the heterotrophic respiration return flux of ^{14}C typically photosynthetically assimilated a decade or two ago when atmospheric $^{14}\text{CO}_2$ was higher than it is today. To be consistent with *Miller et al.* [2012], \hat{C}_{cor} neglects possible point source emissions of $^{14}\text{CO}_2$ from nuclear power plants.

Any particular measurement of C_{ff} is affected by both random and systematic measurement errors. Following *Miller et al.* [2012] and *Lehman et al.* [2013], we estimate the random uncertainty of any C_{ff} measurement as $\hat{\sigma}_{obs,ran} = 1$ ppm. The majority of this uncertainty (0.9 ppm) is due to the random uncertainty in each $\Delta^{14}\text{C}$ measurement.

We assume that each C_{ff} measurement has a bias (systematic error) due to imperfect modeling of backgrounds, imperfect modeling of the transport of emissions associated with heterotrophic respiration, and neglecting the effect of nuclear power plants on C_{ff} . These systematic errors vary from observation to observation. Based on previous analysis [*Miller et al.*, 2012; *Lehman et al.*, 2013], we estimate the standard deviation of the systematic error for C_{ff} measurements, $\hat{\sigma}_{obs,sys}$, to be 0.5 ppm. We estimate the combined uncertainty for any C_{ff} measurement as $\hat{\sigma}_{obs,tot} = \sqrt{\hat{\sigma}_{obs,ran}^2 + \hat{\sigma}_{obs,sys}^2} = 1.1$ ppm.

The high variability of the $C_{ff,obs}$ time series (Figure 2) is expected, in part, because of spatial variation of emissions and their ratios [*Miller et al.*, 2012], temporal variation of wind directions, and associated gas transport, and the variations in altitude where data are acquired. In particular, measured C_{ff} values at approximately 2200 m asl are typically low and have large fractional uncertainties because the altitude of the boundary between the free troposphere and the PBL fluctuates about 2200 m asl. In Figure 3, we show scatterplots of $C_{ff,obs}$ and those surrogate gas enhancements selected by at least one of the four PPR prediction models that we study. Although only CO , NMHCs, and CH_4 are potentially coemitted with CO_2 during fossil fuel combustion, enhancements of these and other surrogate gases are positively correlated with $C_{ff,obs}$ since emissions are strongly correlated over the broad spatiotemporal regions to which our air samples are sensitive [*Miller et al.*, 2012].

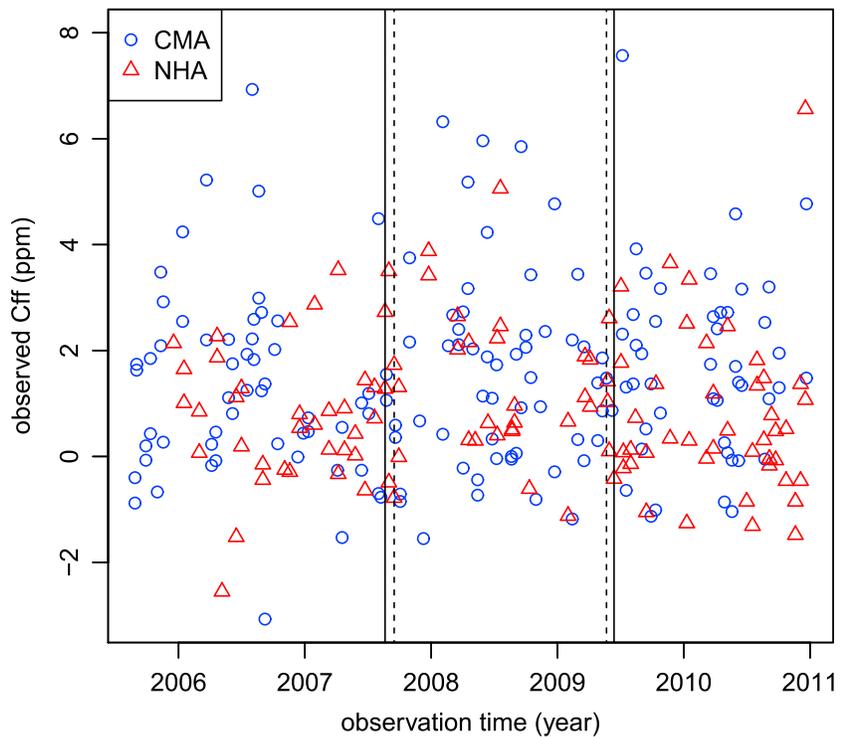


Figure 2. Observed C_{ff} time series. We split the data into a model building part and a test part. The model building part is further split into training and validation data for model selection. For case A, the test data correspond to data acquired before the time associated with the first dashed line. For case B, the test data correspond to data acquired between the two solid lines. For case C, the test data correspond to data acquired after the second dashed line. For case D, every third observation is in the test data.

2.2. C_{ff} Prediction Model

The PPR model prediction of C_{ff} is $C_{ff, \text{pred}}$ where

$$C_{ff, \text{pred}} = \sum_{m=1}^M g_m(w_m \cdot X), \tag{3}$$

where g_m is the m th ridge function, M is the total number of ridge functions, X is a multivariate surrogate gas measurement vector with dimension K , and w_m is a K -dimensional direction vector. Given the mathematical form of the ridge functions, the M direction vectors are estimated from the data. (Logan and Shepp [1975] introduced the term “ridge function” to describe a multivariate function q that maps a n -dimensional real vector X into a real scalar $q(e \cdot X)$ where e is a direction vector in \mathbf{R}^n (n -dimensional Euclidean space) and $e \cdot X$ is the inner product of e and X .) A ridge function varies only along the direction e . In our study, the number of surrogate gases K included in the model, the mathematical form of each ridge function, and the number of ridge functions M are determined by cross validation. In particular, we consider PPR models where the number of ridge functions varies from 1 to 4, and where each ridge function is either a smoothing spline [Craven and Wahba, 1979; Wahba, 1990] with an adjustable effective degrees of freedom (ed.f.) or a supersmoother [Friedman, 1984]. PPR is attractive because of its flexibility; any function can in principle be represented with a PPR model [Diaconis and Shahshahani, 1984]. In contrast, simpler multivariate linear models can not in general represent any function. However, for observed data affected by both random and systematic measurement errors, selecting the form and complexity of the optimal PPR model is nontrivial.

Before constructing a PPR model from any set of model building data, we center and scale measurements for each surrogate gas in the training data by subtracting its sample mean from each measurement and then dividing by the estimated standard deviation of the measurements. The validation data and test data are centered and scaled in exactly the same way.

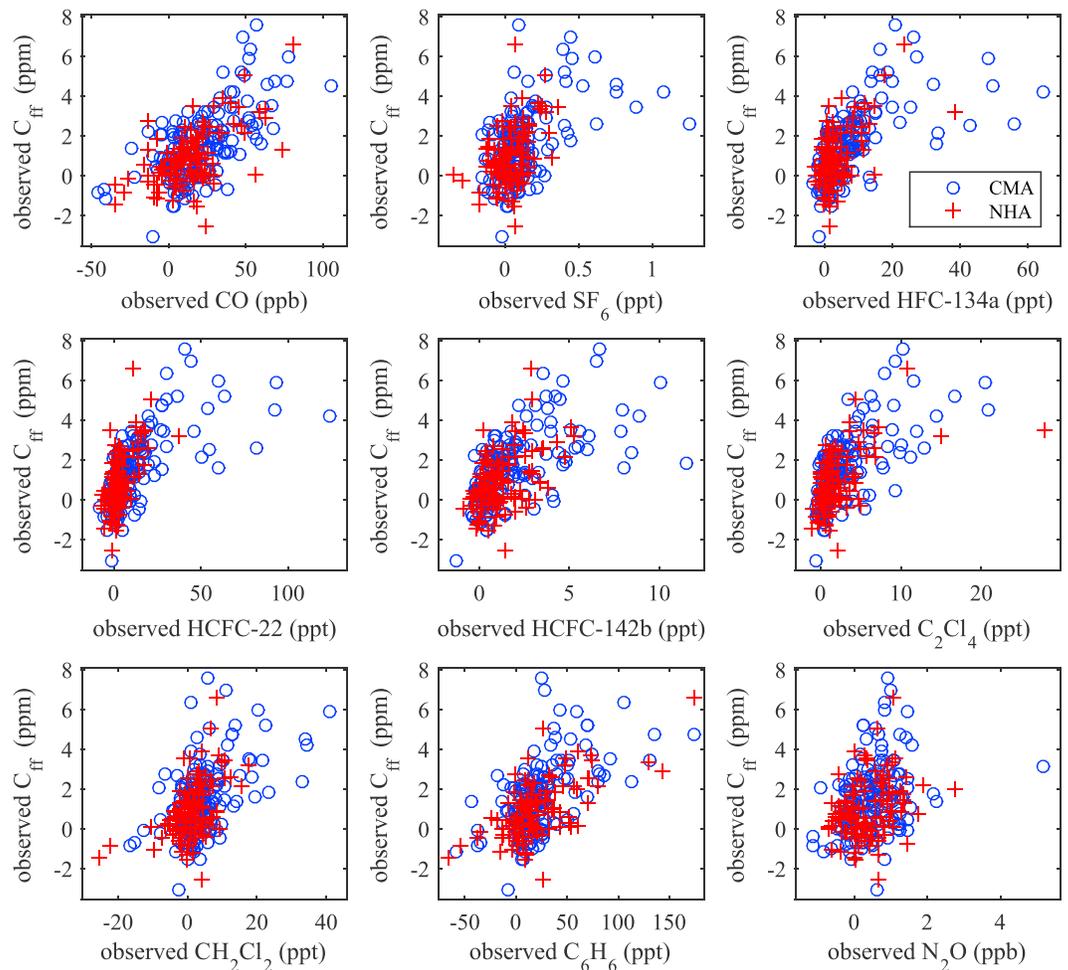


Figure 3. Scatterplots of $\Delta^{14}\text{C}$ -based measurements of C_{fr} and measured surrogate gas enhancements selected for inclusion in a PPR prediction model for at least one of the four cases. Above, ppb and ppt denote mole fractions equal to 10^{-9} and 10^{-12} , respectively.

2.3. Model Selection

Given the number of surrogate gases in the PPR model, we select the optimal combination of the 12 candidate surrogate gases (Table 1) by six-fold cross-validation—an implementation of the K -fold cross-validation method [Hastie et al., 2008]. Every sixth observation of the model building data is assigned to the validation data set. The other observations in the model building data are assigned to the training data set. For each of the six distinct ways to split the model building data into training and validation subsets, we determine a candidate PPR model from the training data and predict C_{fr} values for the validation data set. For each of the six distinct splits of the model building data, we determine the mean-squared difference between measured and predicted C_{fr} for the validation data. The square root of the mean of these six values is the cross-validation statistic (Table 2). We select the identity of the surrogate gases by minimizing this cross-validation statistic (Figure 4). (Since the model building data is different for each of the four cases, variations in both the number and identity of selected surrogate gases as well as the selected ed.f. of the ridge functions in the PPR model are expected (Table 2).) We select the optimal linear multivariate model and the optimal linear single proxy model according to the same cross-validation method described above.

For the data considered here, a PPR model with two ridge functions, where each has a smoothing spline form, works best according to cross-validation. In our study, for each distinct combination of candidate surrogate gases, we vary the effective degrees of freedom (ed.f.) of the ridge functions on a grid. For any particular choice of surrogate gases, the optimal ed.f. value yields the minimum cross-validation statistic. We set the maximum value of ed.f. equal to 3 to encourage smooth ridge functions and impose regularity on the predictions.

Table 2. Selected Surrogate Gases for PPR Prediction Model and Observed RMS Deviation (Observed-Predicted) for Training and Validation Data^a

CASE	RMS Deviation Training Data (ppm)	RMS Deviation Validation Data (ppm)	Selected Gases	Selected ed.f. of Ridge Functions
A	0.927	1.012	(2, 3, 4, 5, 9, 11)	2.82
B	1.027	1.081	(1, 3, 4, 6, 9)	2.95
C	1.079	1.106	(4, 6, 7, 9, 11)	2.66
D	1.051	1.096	(4, 6, 9)	2.90

^aFor each of the four cases, we select a corresponding optimal PPR model by minimizing our cross-validation statistic which is the estimated root-mean-square deviation between predicted and observed values in the validation data (third column in this table).

In our analysis, the number of surrogate gases in a candidate subset varies from two to six. For subsets of size seven or larger, the PPR model had convergence problems. In particular, for a subset with seven or more surrogate gases, the resulting cross-validation statistic was sensitive to the order of surrogate gases in the R code [R Development Core Team, 2013] implementation of the PPR method. However, for subsets of size six or less, the variation of results due to the order of the gases was very slight.

2.4. Analysis

We quantify how well the PPR model determined from model building data predicts test data that are independent of the model building data for four cases (Figure 2). In case A, the test data are the first third of the $C_{ff,obs}$ time series. In case B, the test data are the middle third of the time series. In case C, the test data are the last third of the time series. In case D, the test data include every third observation in the time series. For cases B and D, we make slight adjustments so that the test data have the same number of observations as cases A and C.

For each case, we estimate the uncertainty of the selected model, $\sigma_{pred,post}$ with a bootstrap method [Efron and Tibshirani, 1993] (see Appendix C for details). Neglecting the uncertainty associated with model selection methods leads to overly optimistic uncertainty estimates [Burnham and Anderson, 2002; Claeskens and Hjort, 2008; Hoeting et al., 1999; Hjort and Claeskens, 2003]. Therefore, we estimate an additional component of uncertainty, $\sigma_{pred,model}$ that accounts for imperfection in our model selection method. In particular, our estimate $\hat{\sigma}_{pred,model}$ is the estimated standard deviation of the predictions corresponding to the optimal

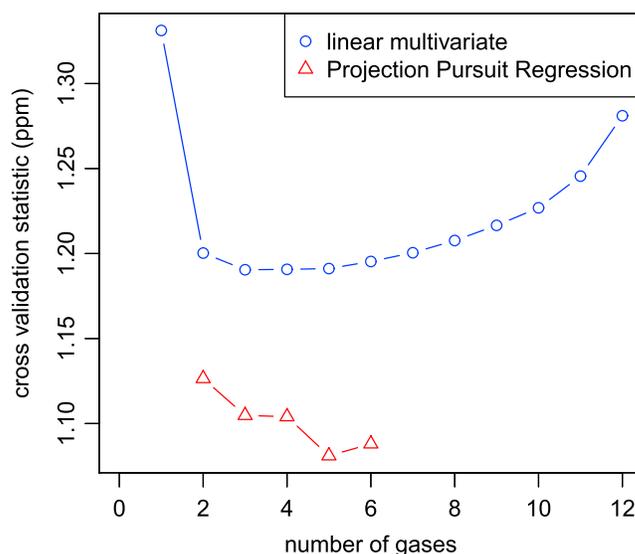


Figure 4. For case B, we show cross-validation statistics corresponding to optimal multivariate linear model and the optimal nonlinear projection pursuit regression models as a function of the number of surrogate gases in the prediction model.

2-gas, 3-gas, 4-gas, 5-gas, and 6-gas PPR models selected by cross validation. Our provisional estimate of the combined uncertainty of the PPR model prediction is

$$\hat{\sigma}_{pred,tot} = \sqrt{\hat{\sigma}_{pred,post}^2 + \hat{\sigma}_{pred,model}^2} \quad (4)$$

Based on equation (4), we estimate the combined uncertainty of any deviation as

$$\hat{\sigma}_{dev,tot} = \sqrt{\hat{\sigma}_{obs,ran}^2 + \hat{\sigma}_{obs,sys}^2 + \hat{\sigma}_{pred,tot}^2} \quad (5)$$

Given $\hat{\sigma}_{dev,tot}$ we test the hypothesis that the expected difference (deviation) between $C_{ff,obs}$ and $C_{ff,pred}$ is 0 ppm based on a chi-square test statistic. To complement the hypothesis test analysis, for each test data set, we estimate a trend (and an associated confidence band) in the deviation ($C_{ff,pred} - C_{ff,obs}$) as a function of $C_{ff,pred}$ with a nonparametric smoothing method.

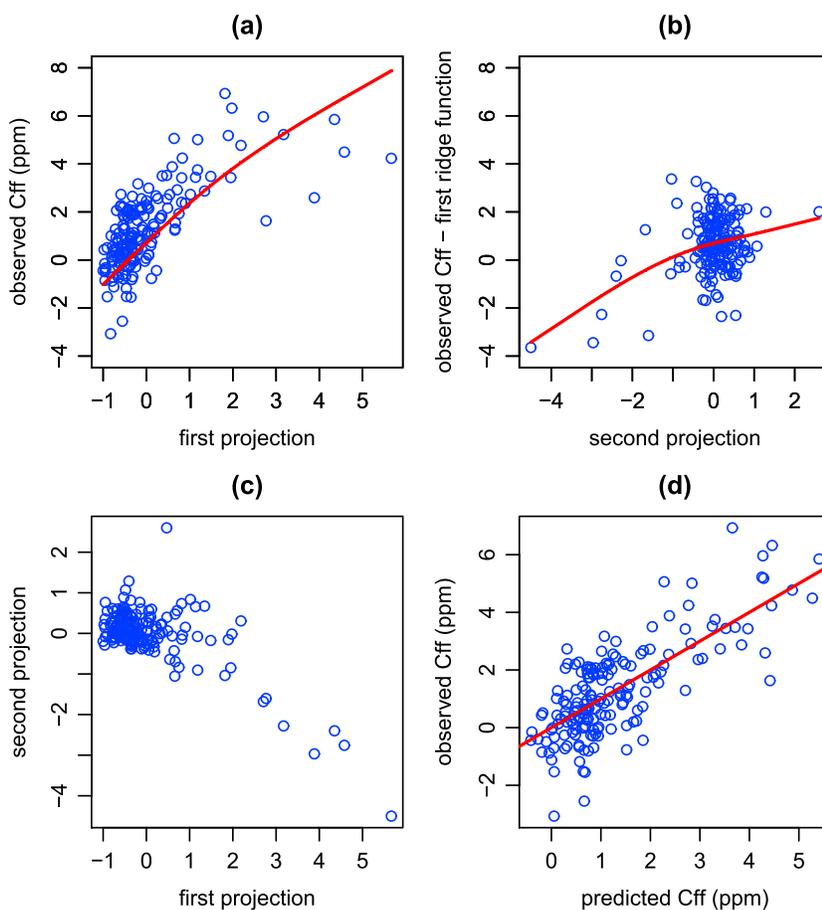


Figure 5. For case C the selected gases for the projection pursuit regression model are HFC-22, C₂Cl₄, CH₂Cl₂, C₆H₆, and N₂O. The direction vectors for the two ridge functions are $w_1 = (0.937, 0.118, -0.152, 0.277, -0.094)$ and $w_2 = (-0.903, 0.311, -0.028, 0.181, 0.232, .)$. (a) First ridge function (solid line) and observed C_{ff}. (b) Second ridge function (solid line) and residual about first ridge function (observed C_{ff} minus first ridge function). (c) Scatterplot of projections of surrogate gas measurements, (d) observed versus predicted C_{ff}, and line of equality. Prediction is sum of the two ridge functions.

To illustrate the PPR approach, we show the two ridge functions for case C (Figure 5) corresponding to the selected PPR model determined from the model building data. Given the first estimated direction vector w_1 and associated ridge function, we form residuals (observation ridge function) and show how well the second ridge function predicts these residuals. The PPR model prediction is the sum of the two ridge functions.

3. Results and Discussion

3.1. Prediction Model Uncertainty

In Figures 6a–6c we show how $\hat{\sigma}_{pred,post}$, $\hat{\sigma}_{pred,model}$, and $\hat{\sigma}_{pred,tot}$ vary with predicted C_{ff} for case D. As discussed later in section 3.3, we expect $\hat{\sigma}_{pred,post}$ to generally increase as C_{ff,pred} increases to large values because random prediction uncertainties for extreme values of C_{ff} are expected to be larger than random prediction uncertainties corresponding to midrange values. The approximate *j*-shaped variation of $\hat{\sigma}_{pred,post}$ with predicted C_{ff} is consistent with this expectation (Figure 6a). More specifically, for C_{ff,pred} greater than 0.5 ppm, $\hat{\sigma}_{pred,post}$ generally increases as C_{ff,pred} increases. For C_{ff,pred} less than 0.5 ppm, $\hat{\sigma}_{pred,post}$ generally increases as C_{ff,pred} decreases. Similar remarks apply to $\hat{\sigma}_{pred,tot}$. For cases A–D, the 0.16, 0.5 and 0.84 the quantiles of the empirical distribution of $\hat{\sigma}_{pred,tot}$ are (0.27, 0.33, 0.49) ppm, (0.14, 0.21, 0.39) ppm, (0.17, 0.22, 0.37) ppm, and (0.15, 0.22, 0.49) ppm. Because of the *j*-shaped dependence of $\hat{\sigma}_{pred,tot}$ on C_{ff,pred}, one should interpret the empirical quantiles of $\hat{\sigma}_{pred,tot}$ determined from all the test data with caution.

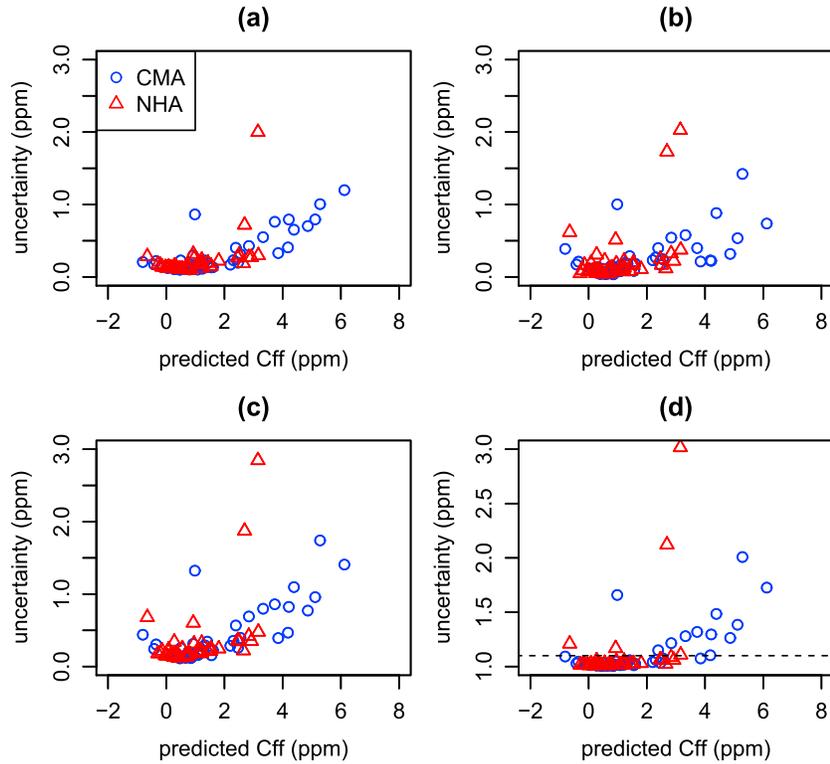


Figure 6. Components and combined components of uncertainty of predicted C_{ff} as a function of predicted C_{ff} for case D. (a) $\hat{\sigma}_{pred,post}$. (b) $\hat{\sigma}_{pred,model}$. (c) $\sqrt{\hat{\sigma}_{pred,post}^2 + \hat{\sigma}_{pred,model}^2}$. (d) $\sqrt{\hat{\sigma}_{pred,post}^2 + \hat{\sigma}_{pred,model}^2 + \hat{\sigma}_{pred,extra}^2}$, where $\hat{\sigma}_{pred,extra}=1$ ppm. The dashed horizontal reference line corresponds to 1.1 ppm. Results for the cases A–C are similar.

3.2. Model Performance

A standard way to quantify prediction model performance is to estimate the square root of the mean-squared (RMS) deviation between observed and predicted values (see Appendix B for mathematical details). The estimated RMS deviation between predicted and observed C_{ff} for all observations in the test data ranges from 1.06 ppm to 1.37 ppm for the four distinct definitions of the test data considered (Table 3). As expected, estimated RMS deviations determined from all observations in the test data are larger than corresponding RMS deviations determined for the model building data. For all four ways of defining the test data, the estimated RMS deviation for the subset $C_{ff,pred} > 2$ ppm is larger than the estimated RMS deviation for the subset $C_{ff,pred} \leq 2$ ppm (Table 3). Uncertainties for individual deviations are also larger, in general, for the $C_{ff,pred} > 2$ ppm subset (Figure 7). As discussed later in section 3.3, one expects the random uncertainty of any prediction model to be higher for extreme C_{ff} values relative to random uncertainties for midrange C_{ff} values. This may explain why estimated RMS deviations are larger for the $C_{ff,pred} > 2$ ppm subset compared to the $C_{ff,pred} \leq 2$ ppm subset. Additionally, physical effects could contribute to inflation of RMS deviations. For instance, RMS deviations may be larger for $C_{ff,pred} > 2$ ppm due to incomplete mixing of local emissions at measurement sites which would create systematic errors in $\Delta_{gas}:C_{ff}$ (Δ_{gas} is shorthand for surrogate gas enhancement) ratios for larger C_{ff} values.

Although a useful performance metric, the RMS deviation statistic does not quantify the uncertainty of the PPR model prediction for any particular observation. Further, the RMS deviation statistic does not inform us if the predicted values are consistent with observations given measurement and prediction uncertainties. To determine if predictions are consistent with observations, we test the null hypothesis that the expected value of the difference between observed and predicted C_{ff} is 0 ppm given our estimates of measurement and prediction uncertainties. For each of the four cases, we determine a chi-square goodness-of-fit statistic

$$\chi_{obs}^2 = \sum_i \frac{(C_{ff,pred}(i) - C_{ff,obs}(i))^2}{\hat{\sigma}_{dev,tot}^2(i)}, \quad (6)$$

Table 3. For the PPR Model, Estimates and Approximate 95% Bootstrap Confidence Intervals for the Root-Mean-Square Value of $C_{ff,pred} - C_{ff,obs}$ for Selected Subsets of Both Model Building and Test Data for Four Cases^a

Subset	Case A	Case B	Case C	Case D
<i>Model Building Data</i>				
$\hat{C}_{ff,PPR} < 2$ ppm	0.87 (0.78, 0.97)	0.99 (0.89, 1.12)	1.04 (0.92, 1.18)	0.98 (0.86, 1.13)
$\hat{C}_{ff,PPR} \geq 2$ ppm	1.11 (0.89, 1.42)	1.18 (0.88, 1.59)	1.28 (0.98, 1.66)	1.31 (1.04, 1.66)
All data	0.93 (0.84, 1.05)	1.03 (0.93, 1.18)	1.09 (0.98, 1.22)	1.06 (0.95, 1.20)
November to February	0.97 (0.78, 1.23)	0.85 (0.68, 1.17)	0.94 (0.77, 1.15)	0.89 (0.70, 1.14)
May to September	0.95 (0.80, 1.14)	1.14 (0.99, 1.35)	1.22 (1.01, 1.45)	1.19 (1.02, 1.41)
<i>Test Data</i>				
$\hat{C}_{ff,PPR} < 2$ ppm	1.25 (1.06, 1.49)	0.94 (0.82, 1.10)	0.98 (0.86, 1.12)	0.95 (0.83, 1.08)
$\hat{C}_{ff,PPR} \geq 2$ ppm	1.73 (1.31, 2.68)	1.39 (1.03, 1.87)	1.41 (0.94, 2.29)	1.51 (1.11, 2.09)
All data	1.37 (1.19, 1.66)	1.06 (0.93, 1.25)	1.07 (0.93, 1.31)	1.12 (0.97, 1.35)
November to February	1.09 (0.90, 1.33)	1.06 (0.76, 1.45)	0.95 (0.63, 1.39)	1.23 (0.75, 2.18)
May to September	1.51 (1.24, 2.03)	1.11 (0.89, 1.44)	1.15 (0.96, 1.48)	1.13 (0.92, 1.36)

^aIn case A, test data are first third of time series. In case B, test data are middle third of time series. In case C, test data are last third of time series. In case D, every third observation is in the test data. Results are in units of $\mu\text{mol/mol}$ (ppm). In this study, $\hat{\sigma}_{pred,extra} = 0$.

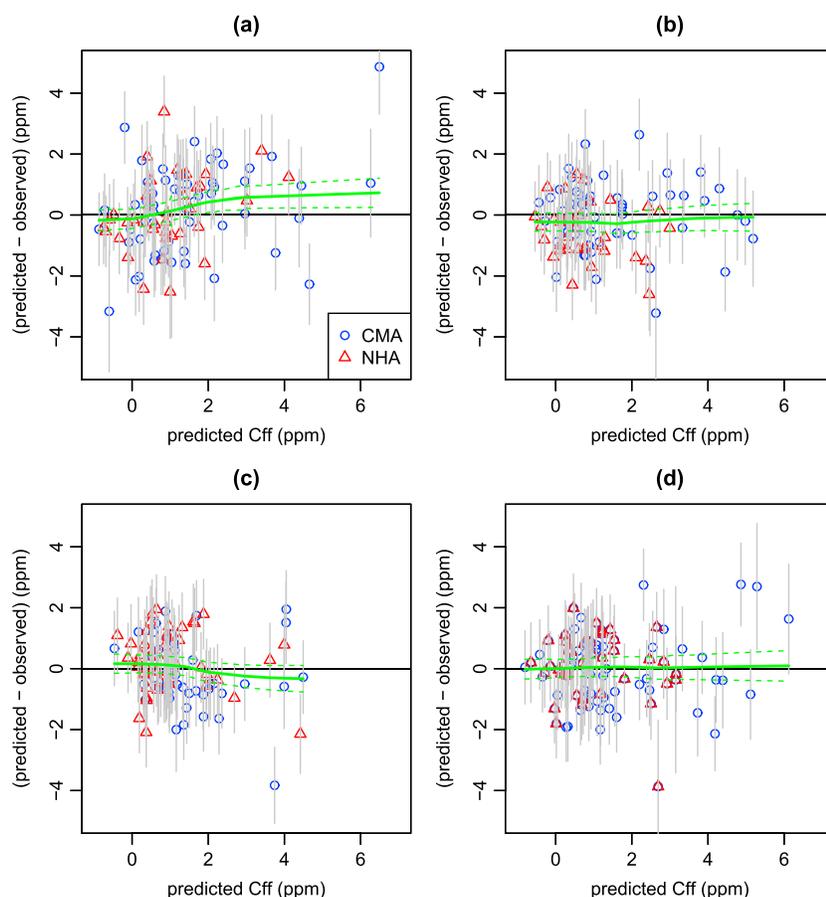


Figure 7. Deviations (predicted-observed) versus predicted C_{ff} for test data. (a) Case A. (b) Case B. (c) Case C. (d) Case D. For each deviation, $C_{ff,pred} - C_{ff,obs}$, we show an associated approximate 68% confidence interval ($C_{ff,pred} - C_{ff,obs} - \hat{\sigma}_{dev,tot}, C_{ff,pred} - C_{ff,obs} + \hat{\sigma}_{dev,tot}$). Solid and dashed lines correspond to trend estimates and associated approximate 95% confidence bands. For cases A–D, the confidence band for the trend estimates includes 0 ppm for the following fraction of observations: 0.71, 1, 1 and 1. Above, $\hat{\sigma}_{pred,extra} = 0$ ppm.

Table 4. Chi-Square Goodness-of-Fit Statistic χ^2_{obs} (Equation (6)), Degrees of Freedom (d.f.), and Associated p -values for Testing the Hypothesis That the Observed and Predicted Test Data Are Consistent

Case	χ^2_{obs}	d.f.	p -value
<i>Cape May, NJ Data</i>			
Case A	71.5	56	0.08
Case B	43.5	58	0.92
Case C	42.3	47	0.67
Case D	46.3	51	0.66
<i>Portsmouth, NH Data</i>			
Case A	38.6	38	0.44
Case B	29.0	36	0.79
Case C	34.7	47	0.91
Case D	27.5	43	0.97
<i>All Data</i>			
Case A	110.1	94	0.12
Case B	72.5	94	0.95
Case C	77.1	94	0.90
Case D	73.7	94	0.94

and an associated p -value (Table 4). When the chi-square test statistic is determined from all 94 observations in each test data set, the p -values for cases A–D are 0.12, 0.95, 0.90, and 0.94, respectively. Since these p -values are high, the evidence for rejecting the null hypothesis is weak.

As a further check of the consistency of the prediction model with observations, we plot trend estimates determined with the LOCFIT method [Loader, 1999, 2010] for the deviation (predicted minus observed) as a function of the predicted value and associated approximate 95% confidence bands (Figure 7). In our implementation of LOCFIT, we fit a local polynomial of degree 0 and set the span parameter to 0.75. The confidence bands are determined by a parametric bootstrap method [Efron and Tibshirani, 1993]. For cases A–D, the fraction of observations for which the approximate 95% confidence bands for the trend include 0 ppm are 0.71, 1, 1, 1.

To study possible variation of model performance with season, we compare predicted and observed C_{ff} values for test data corresponding to (November–February) and (May–September) (Figures 8 and 9). For each case, even though the proxy prediction model is determined from the complete model building data from all seasons, the visual agreement between predicted and observed values for seasonal subsets is good. For the test data, the estimated RMS deviations for the summer subset were, on average, 14% higher than the estimated RMS deviations for the winter subset (Table 3).

3.3. Other Sources of Systematic Uncertainty

In general, if the fractional uncertainties of observations are relatively large and/or the total number of observations is not sufficiently large, hypothesis testing may not reveal real systematic differences between observations and model predictions. Since the uncertainties for $\Delta^{14}\text{C}$ -based C_{ff} measurements analyzed in this work are typically large relative to associated individual measurement values and the range of measured values, more data (particularly measurements corresponding to larger C_{ff} values) may be necessary to produce evidence for possible additional systematic uncertainties not accounted for in our current model. Below, we discuss plausible sources of such additional uncertainty.

As stated earlier, we expect each C_{ff} observation ($C_{\text{ff,obs}}$) to have a nonzero bias (systematic error). Ideally, the unobserved biases of C_{ff} measurements should have a mean value close to 0 ppm and vary independently from observation to observation. Since we have no empirical method to verify these modeling assumptions, it is possible that the mean systematic error of any set of C_{ff} measurements may vary from 0 ppm in a scientifically significant way. To illustrate this effect, suppose that actual systematic errors for any set of measurements are independent realizations of exponentially distributed random variables with theoretical mean and standard deviation of $\hat{\sigma}_{\text{obs,sys}} = 0.5$ ppm. For this case, the expected bias for each measurement is 0.5 ppm.

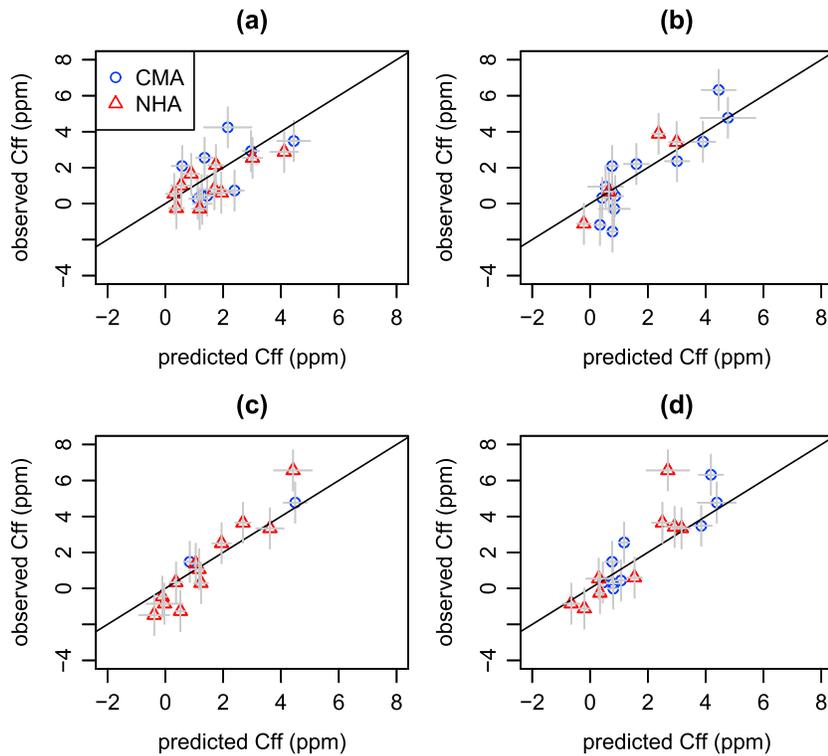


Figure 8. Observed C_{ff} versus proxy model prediction of C_{ff} for subset of test data corresponding to November through February. (a) Case A definition of test data. (b) Case B definition of test data. (c) Case C definition of test data. (d) Case D definition of test data. Line of equality shown in black. RMS deviations between observed C_{ff} and predicted C_{ff} and associated approximate 95% confidence intervals for cases A–D are (in ppm units); 1.09 (0.90, 1.33), 1.06 (0.76, 1.45), 0.95 (0.63, 1.39), and 1.23 (0.75, 2.18). Above, $\hat{\sigma}_{pred,extra} = 0$ ppm.

This would introduce a systematic error of approximately 0.5 ppm into all proxy model predictions. Worse yet, for such an exponential bias model, hypothesis testing could confirm that (biased) observations and (biased) predictions are consistent. Since our goal is to construct a confidence interval for (true) C_{ff} value rather than the associated expected value of a biased measurement of C_{ff} , we must consider the above effect as an additional source of systematic uncertainty. Further, our estimate $\hat{\sigma}_{pred,tot}$ (equation (4)) does not account for systematic uncertainties in measured surrogate gas in the test data due to imperfect background correction, spatial effects due to determining one PPR model from two sites rather one PPR model for each site, systematic temporal variations of surrogate gas emission rates, and systematic effects due to incomplete mixing of emissions at measurement sites.

To account for the systematic effects above, we incorporate an additional component of uncertainty, $\hat{\sigma}_{pred,extra} = 1$ ppm, into our prediction uncertainty budget. Our decision to set $\hat{\sigma}_{pred,extra} = 1$ ppm is based on scientific judgment informed by an additional study (not presented here) of how variability in surrogate gas background estimates affects predicted C_{ff} values. Hence, our final estimate of the combined uncertainty of C_{ff} is $\sqrt{\hat{\sigma}_{pred,post}^2 + \hat{\sigma}_{pred,model}^2 + \hat{\sigma}_{pred,extra}^2}$ (see Figure 6d). We note that an uncertainty budget with $\hat{\sigma}_{pred,extra} = 0$ ppm is sufficient to explain deviations between observed and predicted C_{ff} values according to hypothesis test results (Table 4). Further, any additional systematic uncertainty might vary as a function of $C_{ff,pred}$ (or as a function of other quantities) contrary to our approach.

We note that the combined uncertainty for the prediction of C_{ff} due to systematic effects is $\hat{\sigma}_{pred,sys} = \sqrt{\hat{\sigma}_{pred,model}^2 + \hat{\sigma}_{pred,extra}^2}$ which is greater than $\hat{\sigma}_{obs,sys}$ for all observations. Also, the combined uncertainty for the prediction of C_{ff} due to random effects is $\hat{\sigma}_{pred,post}$ which is generally less than $\hat{\sigma}_{obs,ran}$ (Figure 6a). For case D, for 77 of 94 observations, the combined uncertainties (due to random and systematic effects) for C_{ff} predictions are slightly less than the uncertainty of any $\Delta^{14}C$ -based measurement of C_{ff} (1.1 ppm) (Figure 6d). This phenomenon is possible because the random uncertainty of any model prediction decreases as the number of observations n_{obs} that the model is determined from increases. In our case the PPR model is determined

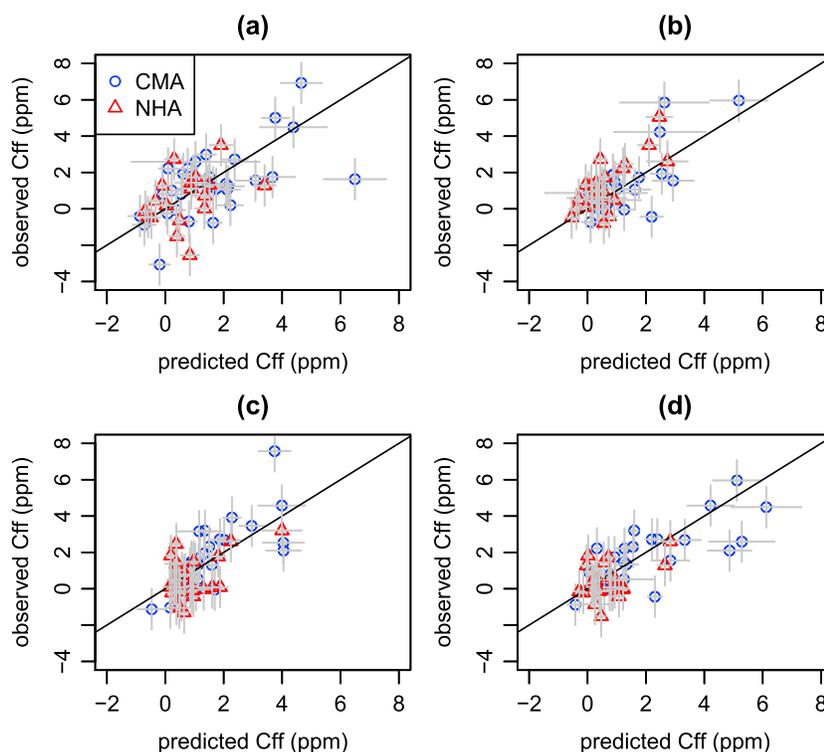


Figure 9. Observed C_{ff} versus proxy model prediction of C_{ff} for subset of test data corresponding to May through September. (a) Case A definition of test data. (b) Case B definition of test data. (c) Case C definition of test data. (d) Case D definition of test data. Line of equality shown in black. RMS deviations between observed C_{ff} and predicted C_{ff} and associated approximate 95% confidence intervals for cases A–D are (in ppm units); 1.51 (1.24, 2.03), 1.11 (0.89, 1.44), 1.15 (0.96, 1.48), and 1.13 (0.92, 1.39). Above, $\hat{\sigma}_{\text{pred,extra}} = 0$ ppm.

from 174 observations where each observation is a vector where one component is observed C_{ff} and the other components are observed surrogate gas enhancements. Hence, for our problem, if the component of prediction uncertainty due to systematic effects is sufficiently small relative to 1.1 ppm and n_{obs} is sufficiently large, and the random uncertainty of surrogate gas enhancements that are input into the PPR prediction model is sufficiently small, the phenomenon is plausible. Moreover, theory predicts this phenomenon for idealized univariate regression models where there are no systematic errors [Mendenhall and Sincich, 1992]. For this idealized case, the width of a confidence interval for the true value (predicted by the fitted regression line) shrinks toward 0 as the number of data points that the regression model is determined from increases without limit. Further, theory predicts wider confidence intervals for extreme observations relative to midrange observations. In general, one expects the random uncertainty of any prediction model (including our PPR proxy model prediction) to be larger for extreme observations relative to midrange observations. In Figure 6a, the approximate j -shaped relationship between the random uncertainty of predicted C_{ff} and predicted C_{ff} may result because the distribution of unobserved C_{ff} is skewed toward lower values.

As discussed earlier, for each of the four cases of interest, we select PPR predictions models based on model building data acquired at both sites. We expect prediction model performance to be best when model building data is acquired at the same site where we predict C_{ff} . Moreover, prediction model performance could vary with the time of surrogate gas measurements and depend on the time period corresponding to when model building data is acquired. Such temporal variations could be diurnal, weekly, seasonal, or long term. Long-term variations are highly likely because of emissions of discontinued refrigerants that change in time, or relative emission rates that vary due to changes in industrial, technological, or energy production processes or regulations. Hence, over sufficiently long time periods, one should recalibrate any surrogate gas prediction model to avoid degradation of model performance. A simple form of recalibration would be to split up the model building data into (nonoverlapping) blocks according to time. One would then determine a PPR model for each block. We note that systematic temporal variability effects could introduce temporal trends in deviations between predicted and observed values of C_{ff} . However, for all four cases, when we fit linear trend

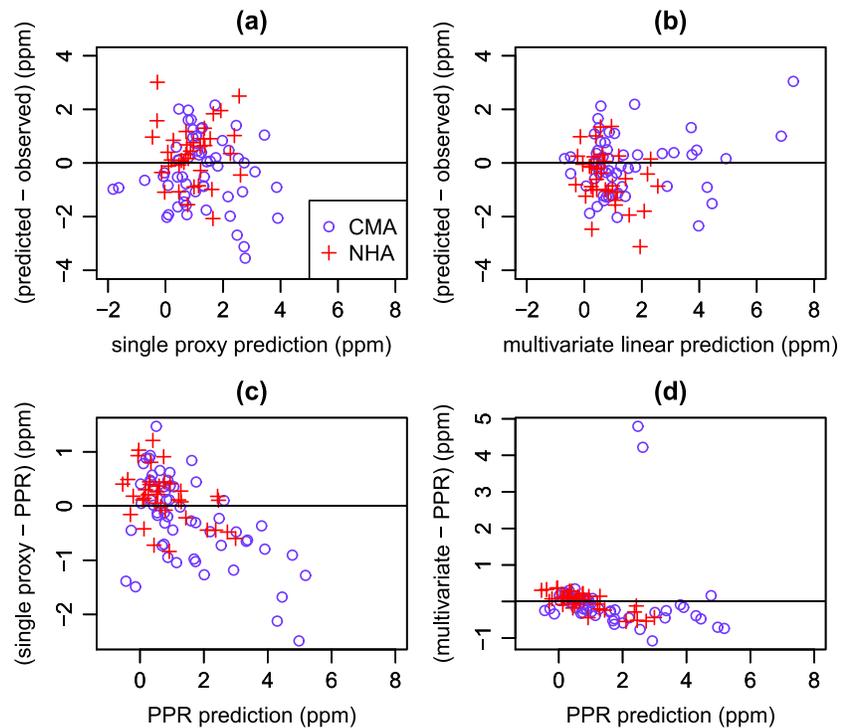


Figure 10. Test data case B. (a) Selected surrogate gases in single proxy prediction model is CO. (b) Multivariate proxy model includes three surrogate gases: HFC-22, C₂Cl₄, and C₆H₆. (c) Difference between single proxy and PPR prediction models. (d) Difference between multivariate and PPR prediction models. Uncertainty intervals not shown.

models to deviations from both sites, the evidence for trends is weak. For cases A–D the estimated slopes and associated uncertainties (in parentheses) are $-0.01(0.21)$ ppm/yr, $0.14(0.18)$ ppm/yr, $0.08(0.22)$ ppm/yr, and $0.02(0.07)$ ppm/yr, respectively. Further, model performance for case C (where the PPR model is extrapolated into the future, and model performance for case D (where the model building data in interwoven with the test data) is similar according to RMS deviation statistics (Table 3) and p -values from hypothesis testing (Table 4).

3.4. Alternative Proxy Models

We also predict C_{ff} with a standard linear multivariate prediction model. For all cases considered, this linear model yields uniformly higher cross-validation statistics compared to the PPR model (see Figure 4 for an example). For case B, where cross validation selects CO as the best surrogate gas for a single proxy linear model, the cross-validation statistics for the optimal single proxy linear model, the optimal multivariate linear model, and the optimal PPR model are 1.33 ppm, 1.19 ppm, and 1.08 ppm, respectively. For this same case, a single proxy model based on CO yields RMS deviation estimates (and approximate 95% confidence interval) of 1.23(1.00, 1.53) ppm and 1.39(1.00, 2.09) ppm for the subsets corresponding to May to September and November to February. In contrast, for the same subsets, the PPR model yields 1.11(0.89, 1.44) ppm and 1.06(0.76, 1.45) ppm.

Although cross-validation statistics serve a critical role for selecting the optimal prediction model, they do not provide a clear picture of how the prediction models vary with respect to one another. To get some insight into this variation, we quantify the difference between the single proxy and multivariate models with respect to the PPR model for the case B test data (Figure 10). The deviation between the single proxy and PPR prediction models is most dramatic for large PPR prediction values. In general, the multivariate model prediction for C_{ff} is less than the PPR model prediction for cases where the PPR prediction is greater than approximately 1 ppm.

In addition to the multivariate linear and PPR prediction models, we also considered a multivariate adaptive regression spline (MARS) model [Friedman, 1991]. According to our cross-validation criterion, the PPR model outperformed the MARS for all cases considered. In some applications, researchers fit PPR models to transformed rather than raw observations. We explored this approach by fitting a PPR model to $\log(\alpha + C_{ff,obs})$

values where $\alpha = 4$ ppm. (We require a positive α because some $C_{ff,obs}$ values are negative.) According to our cross-validation criterion, this approach underperformed the approach presented in this work.

4. Summary

For each of four ways of defining test data, we selected a PPR proxy prediction model by cross validation based on joint analysis of data from NOAA aircraft profiling sites CMA and NHA. The estimated RMS difference between predicted and observed C_{ff} for the test data excluded from the model building process ranged from 1.06 ppm to 1.37 ppm depending on how the test data was defined (Table 3). We also quantified prediction uncertainty for each observation in each test data set (Figure 6). We tested the hypothesis that predicted and observed values are consistent given their associated uncertainties (Table 4). Since evidence of systematic deviations between observed and predicted C_{ff} based on p -values from hypothesis tests is weak, we conclude that development of a scientifically useful multivariate proxy prediction model for C_{ff} is a realistic goal.

In future studies, we plan to expand C_{ff} estimates beyond those based on $\Delta^{14}C$ measurements, by applying our methods to NOAA GGGRN aircraft and tower sites in the U.S. To account for spatial and temporal variations of emission ratios, prediction models would be determined at each site. At each new site, to suppress effects due to systematic temporal variations of surrogate gas emission rates, we currently plan to determine PPR prediction models for contiguous 2 year long blocks of model building data. For sites that lack $\Delta^{14}C$ measurements, development of proxy prediction models based on local model building data is not possible. For such cases, one might predict C_{ff} based on proxy models developed at other sites with similar $\Delta_{gas}:C_{ff}$ ratios. How well such an approach would work is a research question. At many sites, we expect the range of C_{ff} to be larger than for the measurements analyzed in this work. For such sites, on average, we expect lower fractional uncertainties of $\Delta^{14}C$ -based measurements of C_{ff} and predictions of C_{ff} with lower fractional uncertainties.

Additional C_{ff} values determined from surrogate gases ($C_{ff,pred}$) may facilitate a more accurate estimate of U.S. national fossil fuel emissions determined by an atmospheric inverse approach [e.g., Basu *et al.*, 2016]. The utility of proxy model predictions that we plan to acquire at other sites will depend on their yet to be determined uncertainties which we expect to differ from what we determined for the NHA-CMA sites. In future studies, we expect that high-resolution atmospheric transport modeling along with more observations from aircraft vertical profiles and upwind sampling locations at points of continental inflow, such as the Pacific and Gulf coasts, may yield improved determinations of background values for both $\Delta^{14}C$ and surrogate gases. Further, improved representation of heterotrophic respiration and nuclear reactor emissions will also help us reduce and better characterize systematic uncertainties. Because systematic errors can never be eliminated from either C_{ff} or Δ_{gas} measurements, high precision, low bias measurements of $\Delta^{14}C$ remain critical for improving determination of U.S. fossil fuel emissions based on additional proxy model predictions of C_{ff} .

Appendix A

Below, we summarize the theoretical model for C_{ff} and its associated empirical estimate (equation (2)) presented in Miller *et al.* [2012]. Following Turnbull *et al.* [2007], the theoretical CO_2 mole fraction is modeled as due to three sources: background, local fossil fuel emissions, and biospheric sources. This model neglects possible effects due to emissions from nuclear power plants. Given the unobserved true values of C , Δ , C_{bg} , C_{bio} , Δ_{bg} , and Δ_{bio} , one gets the following model equations

$$C = C_{bg} + C_{ff} + C_{bio}, \quad (A1)$$

and

$$\Delta C = \Delta_{bg} C_{bg} + \Delta_{ff} C_{ff} + \Delta_{bio} C_{bio}. \quad (A2)$$

We note that equation (A2) is valid for cases where Δ is small [Vogel *et al.*, 2013]. Since the CMA and NHA sites are far from major sources of fossil fuel CO_2 emissions, we expect negligible systematic error associated with the equation (A2) modeling assumption. We decompose C_{bio} into the sum of a photosynthetic term C_{photo} and a respiratory term C_{resp} . Further, we assume that $\Delta_{photo} = \Delta_{bg}$. Based on equations (A1) and (A2), we estimate C_{ff} as $C_{ff,obs}$ where

$$C_{ff,obs} = \frac{C_{obs}(\Delta_{obs} - \Delta_{bg,obs})}{\Delta_{ff} - \Delta_{bg,obs}} - \frac{\hat{C}_{resp}(\hat{\Delta}_{resp} - \Delta_{bg,obs})}{\Delta_{ff} - \Delta_{bg,obs}}, \quad (A3)$$

and the estimates C_{obs} , Δ_{obs} , $\Delta_{\text{bg,obs}}$ are determined by measurement, and the estimates \hat{C}_{resp} and $\hat{\Delta}_{\text{resp}}$ are determined by a synthesis of experimental measurement and computational modeling. The second term on the right-hand side of equation (A3) is the correction term, \hat{C}_{corr} , that appears in equation (2). This term typically takes values in the range 0.4 ppm to 0.8 ppm during the summer and 0.2 ppm to 0.3 ppm during the winter [Lehman *et al.*, 2013].

Appendix B

In this work, we report estimates of theoretical (true) root-mean-square (RMS) deviations between measured and predicted C_{ff} and associated confidence intervals for the unknown theoretical RMS deviation. Since measured C_{ff} is a function of isotopic ratio measurements, the “true” RMS deviation is undefined if one assumes that the minor and major isotope measurements are realizations of Poisson random variables. This is so because the expected value of the ratio of two Poisson random variables is infinite [Coakley *et al.*, 2005; Coath *et al.*, 2013]. To ensure that true RMS deviation is well defined and reported confidence intervals for C_{ff} are sensible, we restrict analysis to the subsample of C_{ff} measurements where denominator terms in isotopic ratios are positive and $\Delta_{\text{bg,obs}} \neq \Delta_{\text{ff}}$. Our subsampling restriction also ensures that the expected value and theoretical standard deviation of $C_{\text{ff,obs}}$ are well defined. For C_{ff} studies, this subsampling restriction has no practical effect on data acquisition since the probability that a C_{ff} measurement falls outside the subsample where the true RMS deviation is defined is negligible.

Appendix C

The PPR model prediction for C_{ff} values in the test data depends on input surrogate gas measurements in the test data, S_T , and the model building data D_M that determines the PPR model. Each observation in the model building data is a vector of dimension $K+1$ where one component is a $C_{\text{ff,obs}}$ value and the other K components are surrogate gases. Suppose we write the PPR prediction for the i th observation in the test data as

$$C_{\text{ff,pred}}(i) = f(D_M, S_T(i)), \quad (\text{C1})$$

where f denotes the PPR model. To get $\hat{\sigma}_{\text{pred,post}}$ our bootstrap method simultaneously resamples D_M and $S_T(i)$ by nonparametric and parametric bootstrap methods, respectively. The nonparametric bootstrap resampling scheme for D_M accounts for observation-to-observation variation of systematic measurement errors and random uncertainties in both C_{ff} and surrogate gas measurements. The parametric bootstrap resampling scheme for $S_T(i)$ accounts for random uncertainties but not observation-to-observation variation of systematic measurement errors. Our method does not account for the possibility that the mean systematic measurement error over all observations may not be 0 ppm. This additional uncertainty is accounted for with a method described in section 3.3.

Acknowledgments

We thank A. Possolo, A. Pintar, D. Samarov, and J.C. Turnbull for their helpful comments, the NOAA Climate Program Office/AC4 for funding the research that produced the measurements analyzed here, and C. Siso, A. Karion, and others from NOAA and CIRES for technical and logistical contributions related to sampling and analysis of flasks. Contributions to this work by staff of NIST (an agency of the U.S. government) are not subject to copyright in the U.S. and represent an execution of the NIST Greenhouse Gas and Climate Science Measurements Program (Special Programs Office, Associate Director for Laboratory Programs, NIST, U.S. Department of Commerce). Data and software scripts are available in the supporting information.

References

- Anderson, E. C., W. F. Libby, S. Weinhouse, A. F. Reid, A. D. Kirshenbaum, and A. V. Grosse (1947), Natural radiocarbon from cosmic radiation, *Phys. Rev.*, *72*, 931–936.
- Andrews, A. E., et al. (2014), CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: Instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas. Tech.*, *7*, 647–687.
- Arlot, S., and A. Celisse (2010), A survey of cross-validation procedures for model selection, *Stat. Surv.*, *4*, 40–79.
- Ballantyne, A. P., C. B. Alden, J. B. Miller, P. P. Tans, and J. W. C. White (2012), Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years, *Nature*, *488*(7409), 70–72.
- Basu, S., J. B. Miller, and S. Lehman (2016), Separation of biospheric and fossil fuel fluxes of CO₂ by atmospheric inversion of CO₂ and ¹⁴CO₂ measurements: Observation System Simulations, *Atmos. Chem. Phys. Discuss.*, *2016*, 1–34.
- Boden, T. A., G. Marland, and R. J. Andres (2014), *Global, Regional, and National Fossil-Fuel CO₂ Emissions*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., doi:10.3334/CDIAC/00001_V2014.
- Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*, 2nd ed., Springer-Verlag, New York.
- Canadell, J. G., C. Le Quere, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R. A. Houghton, and G. Marland (2007), Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks, *Proc. Natl. Acad. Sci. U.S.A.*, *104*(47), 18,866–18,870, doi:10.1073/pnas.0702737104.
- Claeskens, G., and N. L. Hjort (2008), *Model Selection and Model Averaging*, Cambridge Univ. Press, New York.
- Coakley, K. J., D. S. Simons, and A. M. Leifer (2005), Secondary ion mass spectrometry measurements of isotopic ratios: Correction for time varying count rate, *Int. J. Mass Spectrom.*, *240*(2), 107–120.
- Coath, C. D., R. C. J. Steele, and W. Fred Lunnon (2013), Statistical bias in isotope ratios, *J. Anal. At. Spectrom.*, *28*(1), 52–58.

- Committee on Methods for Estimating Greenhouse Gas Emissions; National Research Council (2010), *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements*, The Natl. Acad. Press, Washington, D. C.
- Craven, P., and G. Wahba (1979), Smoothing noisy data with spline functions, *Numer. Math.*, *31*, 377–403.
- Diaconis, P., and M. Shahshahani (1984), On non-linear functions of linear combinations, *SIAM J. Sci. Stat. Comput.*, *5*(1), 175–191.
- Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap, Monographs and Statistics and Applied Probability 57*, CRC Press, New York.
- Etheridge, D. M., L. P. Steele, R. L. Langenfelds, R. J. Francey, J. M. Barnola, and V. I. Morgan (1996), Natural and anthropogenic changes in atmospheric CO₂ over the last 1000 years from air in Antarctic ice and firn, *J. Geophys. Res.*, *101*(D2), 4115–4128.
- Friedman, J. H. (1984), A variable span smoother, *Rep. LCS 05*, Dep. of Stat. at Stanford Univ., Stanford, Calif.
- Friedman, J. H. (1991), Multivariate adaptive regression splines (with discussion), *Ann. Stat.*, *19*(1), 1–141.
- Friedman, J. H., and W. Stuetzle (1981), Projection pursuit regression, *J. Am. Stat. Assoc.*, *76*(376), 817–823, doi:10.1080/01621459.1981.10477729.
- Graven, H. D., and N. Gruber (2011), Continental-scale enrichment of atmospheric ¹⁴C₂ from the nuclear power industry: Potential impact on the estimation of fossil fuel-derived CO₂, *Atmos. Chem. Phys.*, *11*(23), 12,339–12,349.
- Gurney, K. R., D. L. Mendoza, Y. Zhou, M. L. Fischer, C. C. Miller, S. Geethakumar, and S. de la Rue du Can (2009), High resolution fossil fuel combustion CO₂ emission fluxes for the United States, *Environ. Sci. Technol.*, *43*(14), 5535–5541.
- Hastie, T., R. Tibshirani, and J. H. Friedman (2008), *The Elements of Statistical Learning*, 2nd ed., Springer-Verlag, New York.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, *14*, 382–417.
- Hjort, N. L., and G. Claeskens (2003), Frequentist model average estimators, *J. Am. Stat. Assoc.*, *98*(464), 879–899.
- Knorr, W. (2009), Is the airborne fraction of anthropogenic CO₂ emissions increasing?, *Geophys. Res. Lett.*, *36*, L21710, doi:10.1029/2009GL040613.
- Suess, H. E. (1965), Secular variations of the cosmic-ray-produced carbon 14 in the atmosphere and their interpretations, *J. Geophys. Res.*, *70*(23), 5937–5952.
- Lal, D., and H. E. Suess (1968), The radioactivity of the atmosphere and hydrosphere, *Annu. Rev. Nucl. Part. Sci.*, *18*, 407–434, doi:10.1146/annurev.ns.18.120168.002203.
- Lehman, S. J., J. B. Miller, C. Wolak, J. Southon, P. P. Tans, S. A. Montzka, C. Sweeney, A. Andrews, B. LaFranchi, T. P. Guilderson, and J. C. Turnbull (2013), Allocation of terrestrial carbon sources using ¹⁴C₂: Methods, measurement, and modeling, *Radiocarbon*, *55*(2-3), 1484–1495.
- Levin, I., and U. Karstens (2007), Inferring high-resolution fossil fuel CO₂ records at continental sites from combined ¹⁴C₂ and CO observations, *Tellus B*, *59*(2), 245–250, doi:10.1111/j.1600-0889.2006.00244.x.
- Levin, I., B. Kromer, M. Schmidt, and H. Sartorius (2003), A novel approach for independent budgeting of fossil fuel CO₂ over Europe by ¹⁴C₂ observations, *Geophys. Res. Lett.*, *30*(23), 2194, doi:10.1029/2003GL018477.
- Lingenfelter, R. A. (1963), Production of carbon 14 by cosmic-ray neutrons, *Rev. Geophys.*, *1*(1), 35–55, doi:10.1029/RG001i001p00035.
- Loader, C. (1999), *Local Regression and Likelihood*, Springer, New York.
- Loader, C. (2010), locfit: Local Regression, Likelihood and Density Estimation, R package version 1.5–6 from The Comprehensive R Archive Network (CRAN). [Available at <http://CRAN.R-project.org/package=locfit>, Verified 9 December 2012.]
- Logan, B. F., and L. A. Shepp (1975), Optimal reconstruction of a function from its projections, *Duke Math. J.*, *42*, 645–659.
- Mendenhall, W. M., and T. L. Sincich (1992), *Statistics for Engineering and the Sciences*, Dellen Publishing, San Francisco, Calif.
- Miller, J. B., et al. (2012), Linking emissions of fossil fuel CO₂ and other anthropogenic trace gases using atmospheric ¹⁴C₂, *J. Geophys. Res.*, *117*, D08302, doi:10.1029/2012JD017048.
- R Development Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. [Available at <http://www.R-project.org/>.]
- Stone, C. (1974), Cross-validated choice and assessment of statistical predictions, *J. R. Stat. Soc. B*, *36*(2), 111–147.
- Stone, C. (1977), An asymptotic equivalence of choice of model by cross-validation and Akaike's Criterion, *J. R. Stat. Soc. B*, *39*(1), 44–47.
- Stuiver, M., and H. A. Pollack (1977), Reporting of C-14 data—Discussion, *Radiocarbon*, *19*(3), 365–363.
- Sweeney, C., et al. (2015), Seasonal climatology of CO₂ across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network, *J. Geophys. Res. Atmos.*, *120*, 5155–5190, doi:10.1002/2014jd022591.
- Turnbull, J. C., J. B. Miller, S. J. Lehman, P. P. Tans, R. J. Sparks, and J. Southon (2007), Comparison of ¹⁴C₂, CO, and SF₆ as tracers for recently added fossil fuel CO₂ in the atmosphere and implications for bio-logical CO₂ exchange, *Geophys. Res. Lett.*, *33*, L01817, doi:10.1029/2005GL024213.
- Turnbull, J. C., et al. (2011), Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009, *Atmos. Chem. Phys.*, *11*(2), 705–721, doi:10.5194/acp-11-705-2011.
- Vogel, F. R., S. Hammer, A. Steinhof, B. Kromer, and I. Levin (2010), Implication of weekly and diurnal ¹⁴C calibration on hourly estimates of CO₂-based fossil fuel CO₂ at a moderately polluted site in southwestern Germany, *Tellus B*, *62*(5), 512–520, doi:10.1111/j.1600-0889.2010.00477.x.
- Vogel, F. R., I. Levin, and D. E. J. Worthy (2013), Implications for deriving regional fossil fuel CO₂ estimates from atmospheric observations in a hot spot of nuclear power plant ¹⁴C₂ emissions, *Radiocarbon*, *55*(3–4), doi:10.2458/azu_js_rc.55.16347.
- Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Society for Industrial and Applied Mathematics, Philadelphia, Pa.
- Yim, M.-S., and F. Caron (2006), Life cycle and management of carbon-14 from nuclear power generation, *Prog. Nucl. Energy*, *48*(1), 2–36, doi:10.1016/j.pnucene.2005.04.002.